Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Advanced water safety predictive modelling using trustworthy ANN enhanced with Shapley additive explanations (SHAP) and Pareto analysis

AMRITA DEY

amrita@civil.iitkgp.ac.in

Indian Institute of Technology Kharagpur

VIGNESH M

Vellore Institute of Technology University

Research Article

Keywords: Water safety assessment, Machine learning, Explainable Artificial Intelligence, Shapley additive explanations, Pareto analysis, Prediction models

Posted Date: November 23rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10919658/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Advanced water safety predictive modelling using trustworthy ANN enhanced with Shapley additive explanations (SHAP) and Pareto analysis

Highlights:

- A novel explainable machine learning model predicts water safety with high accuracy.
- ANN model surpasses others in predictive performance.
- SHAP coupled with Pareto analysis highlights critical chemical interactions relevant to water safety evaluation.
- Perchlorate, ammonia, and nitrate significantly affect water safety.

Abstract:

Water safety evaluation and modelling are essential for ensuring sustainability, environmental protection, and freshwater availability. But conventional models often fail to capture the complex, nonlinear relationships among components. This study proposes an integrated framework combining artificial neural networks (ANN) with explainable artificial intelligence techniques, specifically SHAP, along with statistical and decision-making tools such as Pareto analysis, to enhance water safety assessment and prediction. The developed ANN model demonstrated superior predictive capability with high accuracy (R^2 = 0.99, MAD = 5.2892e-6, and RASE = 0.0000123). SHAP analysis was employed to interpret the model predictions, and Pareto analysis was subsequently applied to prioritize the most influential water safety components. The combined analysis revealed that perchlorate, ammonia, and nitrate are the dominant contributors to water safety determination. The findings suggest that the proposed approach offers an interpretable method for water safety prediction, which can benefit water specialists and policymakers.

Keywords- *Water safety assessment; Machine learning; Explainable Artificial Intelligence; Shapley additive explanations; Pareto analysis; Prediction models*

Abbreviations

ANN, Artificial Neural Network; SHAP, SHapley Additive exPlanations; MAD, Mean Absolute Deviation; RASE, Root Absolute Squared Error; R^2, Coefficient of Determination; WSI, water safety indicator

Introduction:

Water is a valuable resource that sustains life and supports the survival of all living species on Earth. Currently the world is facing an immense water crisis due to water pollution, such as water scarcity and degradation of the water ecosystems (Yin et al., 2025). There are several potential sources of water pollution (Nallakaruppan et al., 2024). Water bodies are increasingly polluted by plastics, debris, industrial effluents, domestic waste, and oil spills, leading to severe ecological damage (Zhang et al., 2025). 19 and 23 million tons of plastics reach aquatic ecosystems each year, including rivers, lakes, and seas (Chen et al., 2025, Zhou et al., 2025). Apart from that, due to the increase in pollution and decomposition of waste in the landfills, leachate generation is taking place. These leachates then drain into the soil and, in a way, into the water streams (Wu et al., 2019). The water streams thus get heavily intoxicated with heavy metals and other elements (Shetty et al., 2023). The extreme use of chemical fertilizers and inefficient manure management has worsened the situation (S. Wang et al., 2025). Sustainable Development Goal 6 focuses on clean water and sanitation (O'Dwyer et al., 2024). Water safety is one of the most essential indicators, as it can give information about the condition of the water environment and surrounding ecosystems (Poursaeid, 2024). The indicators give crucial and appropriate information that enhances decision-making processes (Dey & Ashok, 2024). Water safety refers to the overall condition of water in terms of its chemical, physical, and biological characteristics. Therefore, accurately predicting and managing the quality of surface water is of high importance.

Assessing water safety relies heavily on the systematic evaluation of water quality, which requires the collection and analysis of extensive datasets comprising multiple water quality indicators. Several integrative approaches have been proposed, such as the WQI, originally introduced by Horton in 1960. The WQI incorporates ten key water quality parameters considered essential for evaluating the condition and safety of a water body. By aggregating

temporally and spatially varying water quality parameters into a single numerical value, the WQI provides an effective means of interpreting water safety and suitability for various uses. Over time, numerous WQI models have been developed and applied globally, including the United States National Sanitation Foundation Water Quality Index (NSFWQI), the Scottish Research Development Department Water Quality Index (SRDDWQI), the Canadian Council of Ministers of the Environment Water Quality Index (CCMEWQI), the Weighted Arithmetic Water Quality Index (WAWQI), and the British Columbia Water Quality Index (BCWQI) **(D. Kumar et al., 2023),** among others.

There are several conventional assessment techniques that can assess the water quality, but in reality, the chemical components present in the water have highly nonlinear relationships (Nechita et al., 2025). There are complex interactions among different chemical, biological, and environmental factors. Traditional methods often fail to capture these complex relationships (Alizamir et al., 2025).; thus, it limits the prediction of patterns in uncertain conditions (Stanton & Roelich, 2021). This creates a noteworthy research gap in developing trustworthy, data-driven approaches capable of predicting the water safety by learning complex relationships. Therefore, Deep learning (DL), machine learning (ML), and artificial intelligence (AI) are emerging as the potential technologies for prediction and monitoring waste (Song et al., 2023). Water management can be AI-transformed through enhanced predictive accuracy and efficiency (Rashid et al., 2026). Among all advanced AI methods, ANN enhances feature extraction and proves to be a valuable tool in risk assessment (Dey & Ashok, 2025). ANN can effectively handle nonlinear relationships (Cha et al. 2024) and they have demonstrated exceptional predictive accuracy when dealing with even large datasets (Shomope et al., 2024). Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) are used for assessing the performance of the ANN model. The self-learning nature of ANNs allows them to continuously adapt to new data and evolving trends (Li et al., 2025). They can

continuously learn new data and their trends; ANN models can help in the formation of strategic decisions, thus enabling proactive measures to reduce environmental impacts and optimize resource conservation (Maier et al., 2023).

Comprehensive surveys on explainable AI highlight SHAP as one of the key methods for interpreting complex machine learning models due to its solid theoretical foundation in Shapley values and broad applicability across model types (Linardatos et al., 2020). Recent research compares SHAP-based feature selection with traditional importance measures, demonstrating that aggregated SHAP values can effectively rank feature influence across models **(**H. Wang et al., 2024). Therefore, SHAP has been used in diverse prediction contexts (e.g., fault diagnosis, urban environmental modelling), underscoring its versatility in explaining complex machine learning outputs (Santos et al. 2024). Similarly, several studies have adopted Pareto analysis to rank and prioritize features, demonstrating that a limited number of high-impact variables often account for the majority of model predictions, consistent with the Pareto principle. The Pareto principle, often referred to as the 80/20 rule, is widely applied to prioritize key factors where a minority of causes contribute to the majority of the effect, making it a valuable tool for environmental analysis (Chirinza et al., 2024). Similarly, in multi-objective reservoir management and environmental risk assessment, Pareto approaches help focus attention on the most influential components affecting system reliability and risk outcomes (Kong et al., 2021).

Hence, this comprehensive study attempts to achieve the following objectives: (1) to assess and predict water safety in relation to the WSI using ANN and (2) to employ the SHAP method in combination with Pareto analysis to interpret the black-box nature of the machine learning model and to quantify and prioritize the impact of individual chemical components on the corresponding water safety predictions. (3) to enhance the interpretability and transparency of ANN through the incorporation of explanatory artificial intelligence techniques, thereby

strengthening the reliability of the results and supporting informed decision-making in water safety evaluation and management. Through these objectives, the study endeavors to provide a robust, interpretable, and transparent framework for water safety assessment and prediction that can be trusted by both researchers and decision-makers.

2. Materials and Methods:

The approach utilized in this investigation is illustrated in Figure. 1. Starting from section 2.1, the authors have presented an extensive description of the workflow.

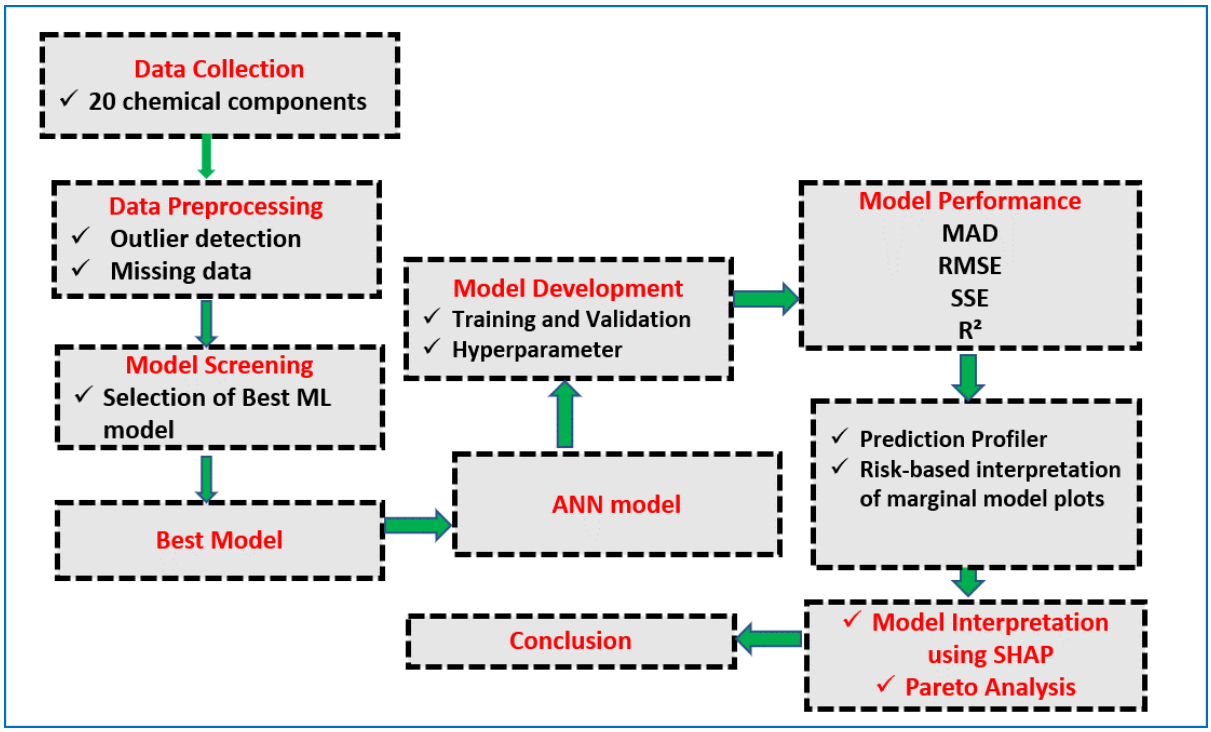

Figure 1: Methodology of the study

2.1 Data Collection:

The dataset was obtained from a publicly available Kaggle repository, originally sourced from internal water quality monitoring records. The dataset is titled Water Quality Analysis (https://www.kaggle.com/datasets/eissaislam/waterquality), and it is composed of 7,999 records and 20 columns for chemical components (various chemicals like aluminum, ammonia, arsenic, barium, cadmium, chloramine, chromium, copper, fluoride, lead, nitrates, nitrites,

mercury, perchlorate, radium, selenium, silver, and uranium). Biological parameters: bacteria and viruses, indicating the presence of biological contaminants. However, we have selected 200 data for this study. The selection of around 200 samples was appropriate for an initial exploratory analysis. The dataset is available in supplementary material sheet 1. This dataset is suitable for environmental studies and machine learning applications and also can be utilized for public safety records and is related to water safety and treatment processes.

2.2 Data Processing:

Although Kaggle is a reputable platform for data scientists, it is important to recognize that the quality and reliability of datasets can often vary. Therefore, the quality of the Kaggle dataset was verified through various steps. First, the source was checked to ensure it was from a credible contributor with positive feedback on the platform. The data scientist was then contacted via a professional networking platform. The dataset was then reviewed for completeness by identifying and addressing missing values or incomplete records. Consistency checks were performed to detect and resolve duplicates and anomalies. In this study, the training data size of the ANN model was 80%, and 20% of the data was used for the validation process.

2.3 Model Screening in Software:

In this work, comprehensive analysis is carried out for the development of water safety prediction models using JMP Pro software involving a wide range of machine learning algorithms, including linear regression, decision trees, support vector machines, and neural networks.

2.4 Best Model:

Through rigorous training and validation, it is noted that neural networks consistently outperformed other models in terms of predictive accuracy and model fit. With their ability to

learn complex patterns and nonlinear relationships from large datasets, ANNs offer a promising approach to predicting water safety.

2.5 Trustworthy ANN model Development:

2.5.1 Neuron model

A neural network has multiple neurons that are connected in multilayer fashion. An algorithm learns to adjust the biases and weights of every layer. The output can be represented mathematically for a neuron as given in equation (1).

$$Yi=f(W_iX+b_i)$$

eq(1)

Where Xi = input value, Wi = weight value, bi = bias, Yi = output, f = activation function

The Tanh activation function has the ability to effectively map input values to a range of −1 to 1. This can contribute towards enhancing the predictive capabilities of the model. In neural networks, the tanh function is commonly used as an activation function. It is defined as follows:

$$Tanh\ (x)=(e^{x}-e^{(-x)})\ /(e^{x}+e^{(-x)}) \qquad \text{eq(2)}$$

Here, x is the input value to the neuron. The *Tanh* function is responsible for mapping input values in the range of −1 to 1. This property can be helpful in certain applications where it is necessary to have outputs centered on 0. The Tanh function is differentiable too, which can be utilized for training neural networks using gradient-based optimization methods.

2.5.2. Model performance Metrics:

Different evaluation metrics, such as the coefficient of determination (R^2), mean absolute deviation (MAD), and root average square error (RASE), that help in assessing the performance of the proposed ANN approach are used. These metrics tell us how well the model is trained and how efficiently it can predict the actual values. The descriptions about the evaluation metrics and related formulas are given below.

Here, $\mathbf{y_i}$: Observed emission values $\mathbf{\hat{y}_i}$**:** Predicted emission values $\mathbf{\bar{y}}$**:** Mean of observed values **n**: Number of observations.

2.5.2.1. Coefficient of Determination (R^2)

The coefficient of determination measures how well the predicted values of neural networks match with the observed WSI based on values of toxic elements present in the water, and it can be calculated using the following formula:

$$R^2 = 1 - (\Sigma (y_i - \hat{y}_i)^2 / \Sigma (y_i - \bar{y})^2)$$

eq (3)

2.5.2.2 Mean Absolute Deviation (MAD)

Mean Absolute Deviation: MAD calculates the average magnitude of errors between predicted and observed values of a neural network. The formula is given by:

$$MAD = (1 / n) \Sigma|y_i - \hat{y}_i| \qquad \text{eq(4)}$$

2.5.2.3 Root Average Square Error (RASE)

Root Average Square Error (RASE) is a metric similar to RMSE that helps in measuring the magnitude of average error, and it is calculated using the following formula:

$$RASE = \sqrt{[\Sigma(y_i - \hat{y}_i)^2 / n]} \qquad \text{eq(5)}$$

R^2 measures how closely predicted values match observed ones; if obtained values are near to 1, it indicates a strong accuracy. The average magnitude of predicted errors is represented by Mean Absolute Deviation (MAD), which provides a simple measure of performance. RASE is sensitive to larger differences as it evaluates average errors by focusing on the squared deviations.

3.1 Proposed ANN-based approaches for water safety assessment

To obtain valuable insights on chemical component trends in water with enhanced interpretability of predicted water safety across the time scales of different countries, a trustworthy ANN approach using SHapley Additive exPlanations (SHAP) is presented in this

work. In this study, twenty chemical components were considered as input variables for prediction, as they were identified as the primary contributors to water toxicity, as illustrated in Fig. 3. The model output is a WSI. Chemical concentrations in water include various chemicals like aluminum, ammonia, arsenic, barium, cadmium, chloramine, chromium, copper, fluoride, lead, nitrates, nitrites, mercury, perchlorate, radium, selenium, silver, and uranium. Based on the concentration of these chemicals, a WSI. Based on the concentrations of these chemical parameters, an Artificial Neural Network (ANN) model was developed to assess water safety. However, the proposed ANN model is also scalable to add more chemical components in the future.

3.1.1 Architecture of ANN

In this work, an ANN model consisting of 20 inputs (aluminum, ammonia, arsenic, barium, cadmium, chloramine, chromium, copper, fluoride, lead, nitrates, nitrites, mercury, perchlorate, radium, selenium, silver, and uranium).

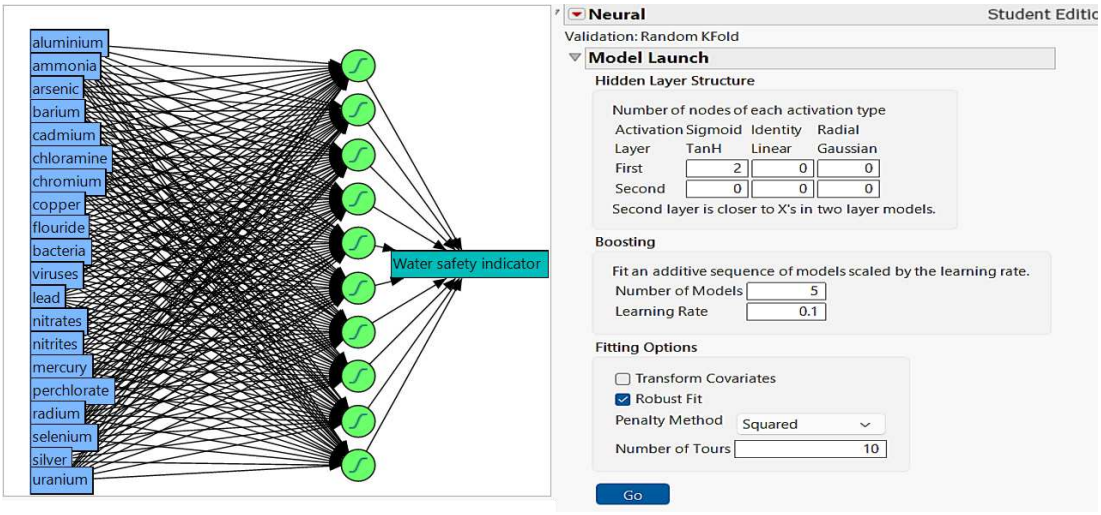

Figure 2: Architecture of ANN model

Biological parameters: Bacteria and viruses, indicating the presence of biological contaminants, and 1 output neuron (WSI) with two-layer architecture is launched as depicted in Fig. 2. The model architecture employed a hyperbolic tangent (tanh) activation function with 2 nodes in the hidden layer. Additionally, a boost of 5 activation nodes was incorporated,

resulting in a total of 10 nodes in the network, which is scaled by a learning rate of 0.1. The number of tours was kept at 10. Here the various chemical components are given as the input, and the WSI values of water are taken as an output of the ANN model. WSI: is_safe (1 (High) for safe, 0 (Low) for not safe), indicating whether the water quality is considered safe for consumption based on the measured parameters. The ANN model was trained with 199 data for capturing the safety/toxicity pattern of the water based on the influence of the chemical components.

3.2 SHapley Additive exPlanations (SHAP)

SHAP (SHapley Additive exPlanations) helps in making more trustworthy models. It is a valuable tool that can be used for understanding the contributions of features in predictions. The SHAP provides a complete explanation of each observation and is in a way different from other existing approaches that do not ensure a uniform distribution of all the predicted features.

The concept of player and game is borrowed from the game theory in SHAP, where it explains the contribution of each feature in the model's prediction. A "player" represents the individual feature. The "game" in SHAP refers to the prediction process of the machine learning model. All the features that are known as players associate together and influence the final prediction. The features are marked as participants in a cooperative game. The SHAP is efficient in calculating even the marginal contributions of all the players involved to every possible coalition, weighted by the probability of those coalitions forming. This thus gives an equitable and fair allocation of the total value of the game among the players.

The SHAP value of player *i* in a game with a set of players *N* and a value function *v* is given by:

$$\phi i(N,v)=\frac{1}{N!}\sum_{S\subset=N\setminus\{i\}}|S|!\,(|N|-|S|-1)!\,[v(S\cup\{i\})-v(S)]$$

eq(7)

Where:

- φ_i(v) is the SHAP value of player i.
- S is a coalition of players that does not include i
- |S| is the size of coalition S
- N is the total number of players
- v(S ∪ {i}) is the value of the coalition S with player i included
- v(S) is the value of the coalition S without player i

This analysis examines the correlation between different elements whose concentration makes the water either safe or unsafe for consumption. SHAP values, a game-theoretic approach, quantify the contribution of each element to a model's prediction. In this context, WSI is developed with respect to different chemicals' presence in water.

3.3 Pareto Analysis:

Pareto analysis is a decision-support technique used to identify and prioritize the most influential factors contributing to a given outcome, based on the Pareto principle (also known as the 80/20 rule). The principle states that a relatively small proportion of causes is often responsible for the majority of the observed effect. In environmental and water safety studies, Pareto analysis is commonly applied to rank parameters according to their relative contribution and to identify critical components that predominantly influence system performance or risk.

Let w_i represent the importance weight of the ith component (e.g., average absolute SHAP value), and let n be the total number of components.

step 1: Normalized Contribution

The percentage contribution of each component is calculated as:

$P_i = (wi / \Sigma wi) \times 100$, where i = 1 to n

eq(8)

Step 2: Cumulative Percentage Contribution

The cumulative contribution is computed by ordering the components in descending order of importance and calculating:

$$Ck = \sum_{i=0}^{k} Pi$$

eq(9)

Interpretation

Components contributing to the cumulative percentage up to a predefined threshold (commonly 80–90%) are considered dominant contributors, while the remaining parameters are regarded as having a relatively minor influence. This approach enables effective prioritization of key parameters, reduces dimensionality, and supports focused monitoring and management strategies in water safety assessment.

4. Results and Discussions:

The developed ANN-based WSI was evaluated using the dataset to examine its predictive performance and interpretability. The following section presents the results obtained from the model, along with a detailed discussion of the influence of individual chemical components on water safety and the insights derived from explainability analyses.

4.1. Evaluation metrics for the ANN Model

The R^2, RASE, and Mean Abs Dev values for training are obtained as 0.9968358, 5.7426e-6, and 5.7426e-6, and for validation they are 0.9999962, 0.0000123, and 5.2892e-6. Moreover, the misclassification rate is also found to be 0. The sum frequency for training was 159, and for validation it is 40.

Therefore, consistent high R^2 and low values of RASE and mean absolute deviation show that the proposed ANN model can be a valuable tool for analysing potential water safety based on the different chemical compositions present in it.

4.2 Statistical Performance Measures

The statistical performance metrics presented in Table 2 indicate the excellent classification capability of the developed ANN model for predicting the water safety indicator. The model achieved 25 true positive (TP) and 15 true negative (TN) predictions, with zero false positive (FP) and false negative (FN) cases, demonstrating error-free classification. The sensitivity value of 1.000 indicates that the model successfully identified all actual positive cases, while the specificity value of 1.000 confirms that all negative cases were correctly classified. Similarly, the precision value of 1 indicates that every predicted positive case was accurate.

The overall accuracy of the model was also 1, signifying 100% correct classification. Furthermore, the F1 score, which represents the balance between precision and recall, was found to be 1, indicating perfect model consistency. The Matthews Correlation Coefficient (MCC), which measures the overall quality of binary classification, also achieved a value of 1, confirming that the model exhibits outstanding predictive reliability and strong agreement between predicted and actual classifications.

4.3 Confusion Matrix and Confusion Rates

A confusion matrix is a performance evaluation tool used to assess the classification ability of a predictive model. It compares the actual class values with the predicted class values and provides detailed information about correct and incorrect predictions. In the present study, the confusion matrix was developed to evaluate the performance of the artificial neural network (ANN) model in classifying the water safety indicator into low and high categories for both training and validation datasets.

4.3.1 Confusion Matrix for Training Dataset

The confusion matrix for the training dataset shows that the ANN model correctly classified all observations. Out of 58 actual low water safety samples, all 58 were correctly predicted as

low, while none were misclassified. Similarly, all 101 high water safety samples were accurately predicted as high, with zero misclassification.

The confusion rates further confirm the model's accuracy. The prediction rate for low water safety samples was 1.000, indicating 100% correct classification, while the misclassification rate was 0.000. Likewise, high water safety samples were also predicted with complete accuracy, demonstrating excellent model learning during training.

4.3.2 Confusion Matrix for Validation Dataset

The confusion matrix for the validation dataset, as shown in Table 1, also demonstrates perfect classification performance. Among the validation samples, all 15 low water safety samples were correctly identified, and all 25 high water safety samples were accurately predicted. No incorrect classification was observed in either category. The confusion rates indicate that the model achieved a prediction accuracy of 1.000 for both low and high-water safety indicators, confirming the strong generalization capability of the model when applied to unseen data.

Table 1. Confusion Matrix and Confusion Rates for Validation

Confusion Matrix			**Confusion Rates**	
Actual	**Predicted Count**		**Actual**	**Predicted**
Low	15	0	**Water Safety Indicator**	**Low**
High	0	25	Low	1.000
			High	0.000

The results of the confusion matrix and confusion rates indicate that the developed ANN model possesses exceptional classification capability for predicting water safety indicators. The absence of false positive and false negative predictions suggests that the selected input parameters strongly influence water safety classification. Furthermore, the consistency of results across both training and validation datasets indicates that the model is stable and reliable for predicting water safety conditions.

4.4 Lift Curve:

Figure 3 depicts the lift curve, indicating strong discriminative capability of the neural network model, particularly in identifying low water safety conditions, with lift values exceeding 2.5 in early ranked portions of data, demonstrating significant improvement over random classification.

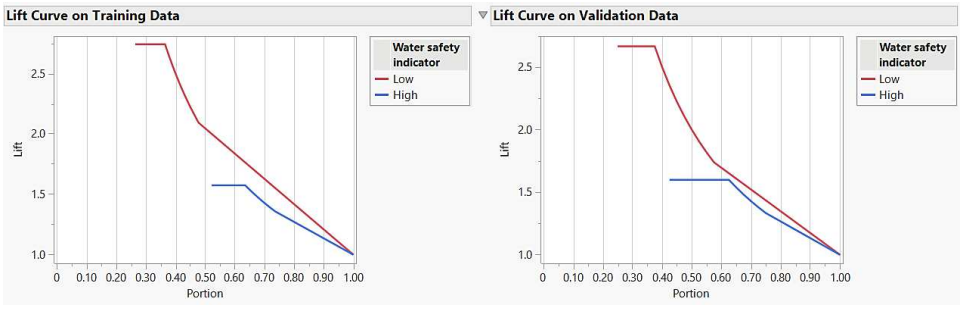

Figure 3: Lift Curve

4.5 ROC curve

In the present study, the ROC curves in Fig. 4 for both low and high-water safety indicators lie near the upper-left corner of the plot, indicating superior classification performance. The Area Under the Curve (AUC) values for both training and validation datasets were found to be 1.000 for low and high-water safety categories, signifying perfect discrimination between the classes. These results confirm that the developed ANN model can accurately differentiate between low- and high-water safety conditions without misclassification.

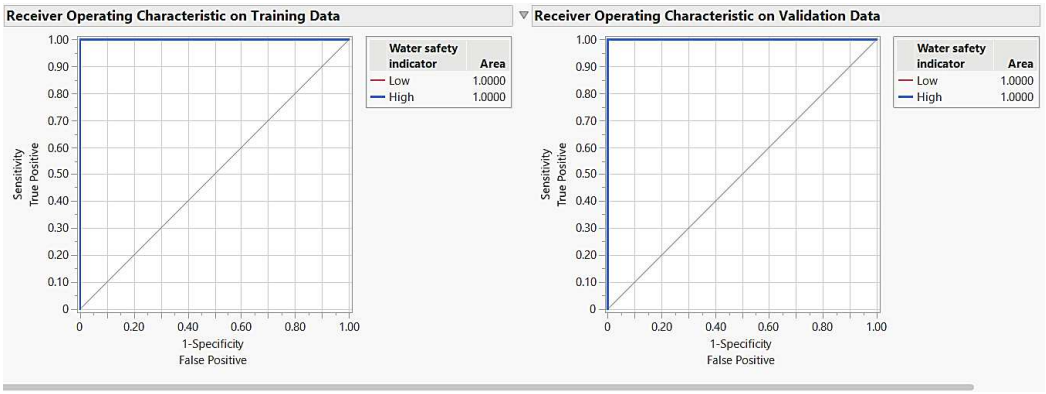

Figure 4: ROC Curve

Furthermore, the similarity in ROC performance between training and validation datasets indicates strong model stability and generalization capability, suggesting that the model is reliable for water safety classification and prediction.

4.6. Interpretation of model predictions using SHAP

The SHAP explanation was employed to analyze the underlying process of the outcomes provided by the ANN model, which exhibited superior performance in comparison to other ML models. Fig. 5 depicts the mean absolute SHAP values (i.e., feature importance) of the ANN model.

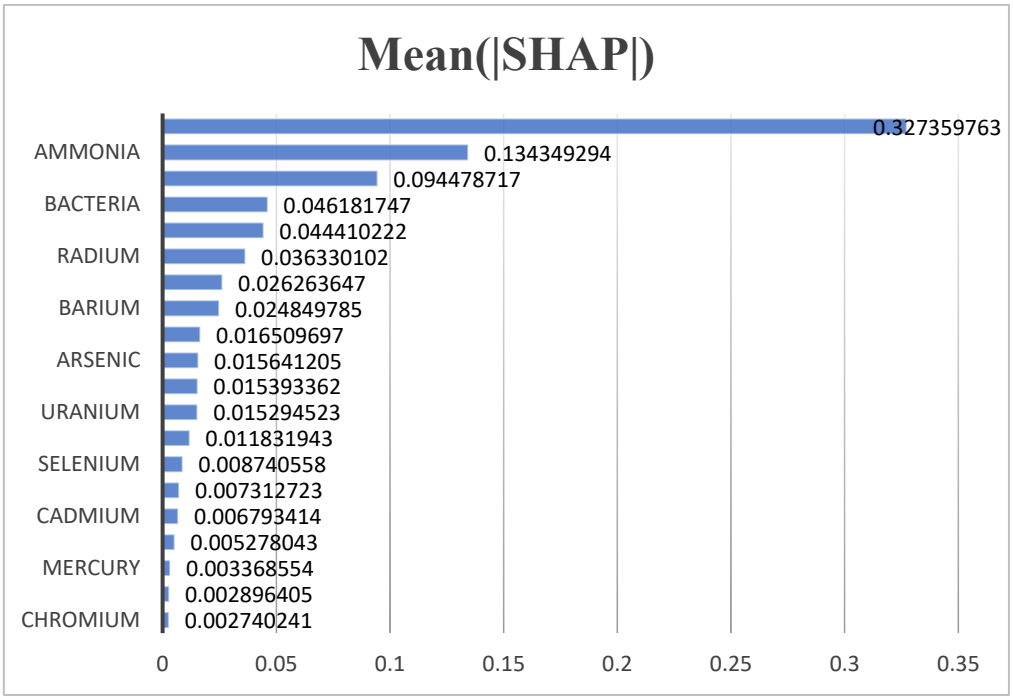

Figure 5: Feature importance ranking

Based on the feature importance in relation to the overall model predictions, perchlorate exhibited the highest mean SHAP value of 0.327, indicating that it is the most important contributing element in predicting WSI. Ammonia ranked second in the list, with a mean SHAP value of 0.134, whereas nitrates (0.094), bacteria (0.046), chloramine (0.044), and radium (0.036) had a nearly equal impact on the model predictions. The effect of virus (0.026), barium (0.024), aluminum (0.016), and arsenic (0.015) on the overall output was quite minimal compared to other water quality parameters.

4.7 SHAP Summary Plot

As presented in Fig. 6, the SHAP summary plot illustrates the relationship between the contribution of each feature value (0-1) and its magnitude; here high is red and low is blue. There is a color bar on the right side of the vertical axis that shows the magnitude of the feature's eigenvalue, where each row represents the SHAP value of the contribution of each

feature to the model output. ANN determined that the parameter perchlorate has the greatest impact on the WSI, and lower values of perchlorate led to an increase in WSI values and vice versa. In a nutshell, a low perchlorate leads to a high WSI, indicating excellent drinking water quality.

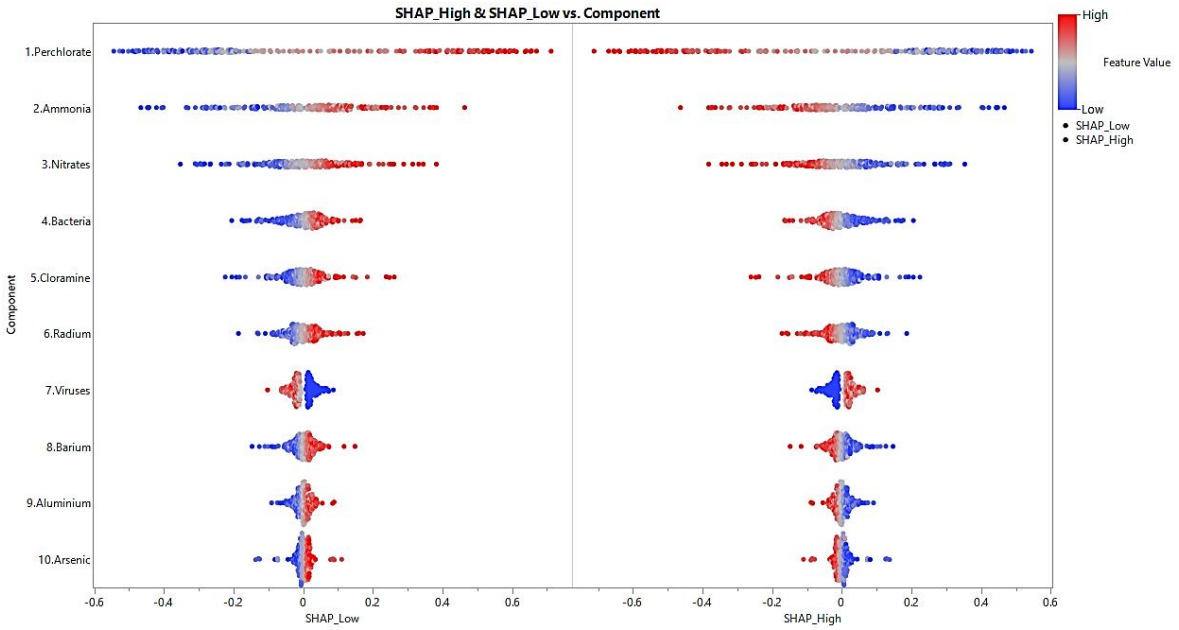

Figure 6: Scatter diagram of the comprehensive influence of the factors on water safety

Similar to the perchlorate, ammonia and nitrate had negatively influenced the target. When ammonia and nitrate are minimal, WSI increases and vice versa. Similarly, high bacteria, chloramine, radium, viruses, barium, aluminum, and arsenic values corresponded to lower WSI values, and vice versa. This indicates that high bacteria and virus concentrations are unsuitable for drinking water, as they lead to low WSI values. Interestingly, SHAP effectively elucidates the precise correlations between chemical components and overall water quality while referencing the standard scientific principles in the field. For instance, Sánchez et al. (2007) also performed a strong relationship study between the WSI and the dissolved components, and SHAP has effectively identified and accurately depicted this correlation. The SHAP values of all the chemical contaminants are presented in the supplementary material.

4.8 Pareto Analysis

Figure 7 presents the Pareto chart illustrating the relative contribution of individual chemical components to water safety based on their average absolute SHAP values. The bars represent the individual influence of each component, while the cumulative percentage curve indicates their combined contribution to overall water safety determination. The components are arranged in descending order of importance, highlighting the disproportionate influence of a limited number of parameters.

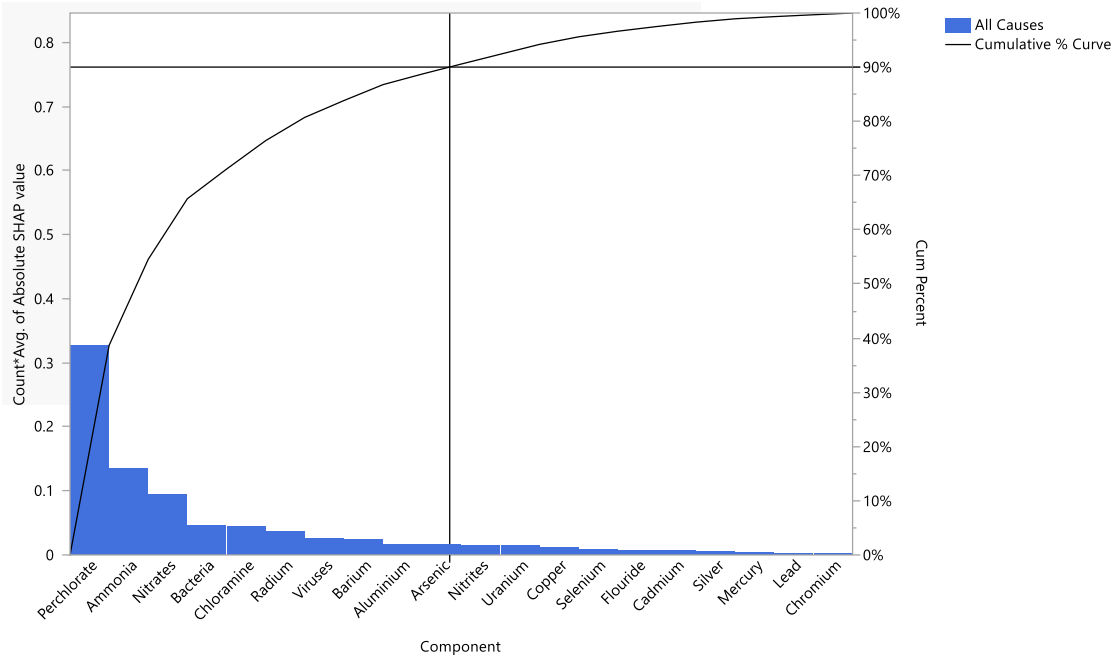

Figure 7: Pareto chart

The cumulative contribution curve shows that the first eight components, perchlorate, ammonia, nitrates, bacteria, chloramine, radium, viruses, barium, aluminum, and arsenic account for approximately 90% of the total cumulative influence on water safety. This observation aligns with the Pareto principle, demonstrating that a relatively small number of key parameters govern the majority of water safety outcomes. In contrast, the remaining components contribute marginally and have a limited effect on the model output. These

findings emphasize the importance of prioritizing high-impact chemical and microbiological parameters in water safety monitoring, risk assessment, and management strategies.

4.6.1 Relative Weights and Contribution of Water Safety Components

Pareto analysis was performed to identify the most influential water quality parameters contributing to the water safety indicator using the average absolute SHAP (Shapley Additive Explanation) values derived from the developed predictive model. The results presented in Table 2 rank the parameters based on their relative contribution to the model output, enabling prioritization of critical contaminants affecting water safety. The analysis revealed that perchlorate exhibited the highest contribution, accounting for approximately 39% of the total influence on the water safety indicator. Ammonia and nitrates were identified as the next most significant parameters, increasing the cumulative contribution to 55% and 66%, respectively. Bacteria and chloramine further contributed to the cumulative influence, collectively raising the total contribution to approximately 76%. These findings indicate that a small number of parameters predominantly influence water safety classification, demonstrating the applicability of the Pareto principle in environmental risk assessment.

Table 2: Per Unit Rates of Components

Cause	Weight (Avg. of Absolute SHAP values)	Cumulative Rate %	Rate
Perchlorate	0.327359763	39%	0.386939
Ammonia	0.134349294	55%	0.158801
Nitrates	0.094478717	66%	0.111674
Bacteria	0.046181747	71%	0.054587
Chloramine	0.044410222	76%	0.052493
Radium	0.036330102	81%	0.042942
Viruses	0.026263647	84%	0.031044
Barium	0.024849785	87%	0.029372
Aluminium	0.016509697	89%	0.019514
Arsenic	0.015641205	91%	0.018488
Nitrites	0.015393362	92%	0.018195
Uranium	0.015294523	94%	0.018078
Copper	0.011831943	96%	0.013985
Selenium	0.008740558	97%	0.010331
Flouride	0.007312723	98%	0.008644
Cadmium	0.006793414	98%	0.008030
Silver	0.005278043	99%	0.006239
Mercury	0.003368554	99%	0.003982
Lead	0.002896405	100%	0.003424

Cause	Weight (Avg. of Absolute SHAP values)	Cumulative Rate %	Rate
Chromium	0.002740241	100%	0.003239
Pooled Total	0.846023945	100%	0.050000

Moderately influential parameters such as radium, viruses, barium, aluminum, and arsenic contributed incrementally, collectively increasing the cumulative contribution beyond 90%. The remaining parameters, including nitrites, uranium, copper, selenium, fluoride, cadmium, silver, mercury, lead, and chromium, exhibited comparatively lower individual contributions but still played a supporting role in the overall prediction accuracy of the model.

The Pareto analysis highlights that prioritizing monitoring and control of high-impact contaminants such as perchlorate, ammonia, nitrates, bacterial contamination, and chloramine could significantly improve water safety management strategies. Furthermore, the results emphasize the effectiveness of SHAP-based interpretability in identifying dominant parameters influencing water quality, thereby providing valuable insights for sustainable water resource management and targeted regulatory interventions.

5. Conclusions:

The primary objective of this study was to assess and predict the water safety in respect to the utilization of a trustworthy ANN model. The study also utilized explainable AI, specifically the SHAP method and Pareto analysis, to interpret the results. Initially, the WSI was determined based on 20 chemical components: aluminium, ammonia, arsenic, barium, cadmium, chloramine, chromium, copper, fluoride, lead, nitrates, nitrites, mercury, perchlorate, radium, selenium, silver, and uranium. Subsequently, ANN was developed using chemical components as the features, with WSI serving as the target variable. The models' performance was later assessed using four statistical indices, namely MAD, RMSE, and R^2.The developed ANN exhibited excellent predictive capability, achieving a high coefficient of determination (R^2 = 0.99) and low error values, including a MAD = 5.2892e-6 and RASE = 0.0000123. The results demonstrated that the ANN model outperformed conventional

approaches, highlighting the strong applicability of machine learning techniques in water safety assessment. To further enhance interpretability, SHAP-based global explanations were used to analyse the outputs of the ANN model. The analysis revealed that perchlorate, ammonia, and nitrate exert the strongest influence on the WSI, while biological parameters such as bacteria, along with chloramine, radium, viruses, and barium showed a comparatively moderate impact. Alumnium and arsenic were found to consistently influence the predictions, with relatively lower contributions. Furthermore, Pareto analysis was applied to the SHAP-derived feature importance values to prioritize the most influential water safety components. The results confirmed that a limited subset of parameters accounts for the majority of the variation in water safety predictions, supporting the Pareto principle and enabling effective parameter prioritization.

This integrated ANN-SHAP-Pareto framework not only improves prediction accuracy but also provides actionable insights by identifying critical parameters that ensure focused monitoring and management. Overall, this study advances water safety assessment by combining predictive accuracy with interpretability, offering a trustworthy and transparent decision-support tool for water professionals and policymakers. The insights derived from SHAP and Pareto analyses enhance understanding of feature importance, interrelationships among parameters, and their relative contributions, thereby supporting informed and justified decision-making in sustainable water resource management.

References:

Abdel-Shafy, H. I., Ibrahim, A. M., Al-Sulaiman, A. M., & Okasha, R. A. (2023). Landfill leachate: Sources, nature, organic composition, and treatment: An environmental overview. *Ain Shams Engineering Journal*, *15*(1), 102293. https://doi.org/10.1016/j.asej.2023.102293

Alizamir, M., Heddam, S., & Kim, S. (2025). A robust and explainable deep learning model based on an LSTM-CNN framework for reliable FDOM prediction in water quality monitoring: Incorporating

SHAP analysis for enhanced interpretability. *Process Safety and Environmental Protection*, *201*, 107594. https://doi.org/10.1016/j.psep.2025.107594

Cha, G., Park, C., & Kim, Y. (2024). Optimal machine learning model to predict demolition waste generation for a circular economy. *Sustainability*, *16*(16), 7064. https://doi.org/10.3390/su16167064

Chen, B., Zhou, Y., Hossain, K. B., Chen, Y., Zhu, X., Li, H., Xu, X., & Cai, M. (2025). Microplastic in tropical island estuaries in China: Source identification and management framework development. *Marine Environmental Research*, *213*, 107643. https://doi.org/10.1016/j.marenvres.2025.107643

Chirinza, N., Muguirrima, P., Zerpa, F. a. L., Martin, A. R., & Pino, C. a. M. (2024). Resources optimization using Pareto analysis for sea water desalination plants. *Cleaner Water*, *2*, 100054. https://doi.org/10.1016/j.clwat.2024.100054

Dey, A., & Ashok, S. D. (2024). Fuzzy logic based qualitative indicators for promoting extended producer responsibility and sustainable food packaging waste management. *Environmental and Sustainability Indicators*, *24*, 100534. https://doi.org/10.1016/j.indic.2024.100534

Dey, A., & Ashok, S. D. (2025). Trustworthy and Human Centric neural network approaches for prediction of landfill methane emission and sustainable waste management practices. *Waste Management*, *195*, 44–54. https://doi.org/10.1016/j.wasman.2025.01.017

https://www.kaggle.com/datasets/eissaislam/waterquality

Li, Z., Zhou, Z., Wang, H., Li, X., Shi, X., Xiao, J., Yang, Z., Sun, M., Li, X., & Jia, H. (2025). Artificial intelligence-incorporated prediction for urban flooding processes in the past 20 years: A critical review. *Environmental Modelling & Software*, *192*, 106525. https://doi.org/10.1016/j.envsoft.2025.106525

Linardatos, P., Papastefanopoulos, V., & Kotsiantis, S. (2020). Explainable AI: A review of Machine Learning Interpretability Methods. *Entropy*, *23*(1), 18. https://doi.org/10.3390/e23010018

Kong, Y., Mei, Y., Wang, X., & Ben, Y. (2021). Solution Selection from a Pareto Optimal Set of Multi-Objective Reservoir Operation via Clustering Operation Processes and Objective Values. *Water*, *13*(8), 1046. https://doi.org/10.3390/w13081046

Kumar, D., Kumar, R., Sharma, M., Awasthi, A., & Kumar, M. (2023). Global water quality indices: Development, implications, and limitations. *Total Environment Advances*, *9*, 200095. https://doi.org/10.1016/j.teadva.2023.200095

Maier, H. R., Galelli, S., Razavi, S., Castelletti, A., Rizzoli, A., Athanasiadis, I. N., Sànchez-Marrè, M., Acutis, M., Wu, W., & Humphrey, G. B. (2023). Exploding the myths: An introduction to artificial neural networks for prediction and forecasting. *Environmental Modelling & Software*, *167*, 105776. https://doi.org/10.1016/j.envsoft.2023.105776

Nallakaruppan, M. K., Gangadevi, E., Shri, M. L., Balusamy, B., Bhattacharya, S., & Selvarajan, S. (2024). Reliable water quality prediction and parametric analysis using explainable AI models. *Scientific Reports*, *14*(1), 7520. https://doi.org/10.1038/s41598-024-56775-y

Nechita, C., Iordache, A. M., Roba, C., Apetri, A. N., & Camarero, J. (2025). Non-linear relationships between climate and toxic metals in stressed polluted areas. *Environmental Research*, *287*, 123124. https://doi.org/10.1016/j.envres.2025.123124

O'Dwyer, A. (2024). Blue gold: Clean water and sanitation (SDG 6). In *Teaching the Sustainable Development Goals to Young Citizens (10-16 years)* (pp. 194-214). Routledge.

Poursaeid, M. (2024). Comprehensive water quality indicators modeling by environmental protection view using multi optimized weighted ensemble machine learnings. *Process Safety and Environmental Protection*, *193*, 696–709. https://doi.org/10.1016/j.psep.2024.11.042

Sánchez, E., Colmenarejo, M. F., Vicente, J., Rubio, A., García, M. G., Travieso, L., & Borja, R. (2007). Use of the water quality index and dissolved oxygen deficit as simple indicators of watersheds pollution. *Ecological Indicators*, *7*(2), 315–328. https://doi.org/10.1016/j.ecolind.2006.02.005

Santos, M. R., Guedes, A., & Sanchez-Gendriz, I. (2024). SHapley Additive ExPlanations (SHAP) for efficient feature selection in rolling bearing fault diagnosis. *Machine Learning and Knowledge Extraction*, *6*(1), 316–341. https://doi.org/10.3390/make6010016

Shetty, S. S., D, D., S, H., Sonkusare, S., Naik, P. B., N, S. K., & Madhyastha, H. (2023). Environmental pollutants and their effects on human health. *Heliyon*, *9*(9), e19496. https://doi.org/10.1016/j.heliyon.2023.e19496

Shomope, I., Tawalbeh, M., Al-Othman, A., & Almomani, F. (2024). Predicting biohydrogen production from dark fermentation of organic waste biomass using multilayer perceptron artificial neural network (MLP–ANN). *Computers & Chemical Engineering*, *192*, 108900. https://doi.org/10.1016/j.compchemeng.2024.108900

Stanton, M. C. B., & Roelich, K. (2021). Decision making under deep uncertainties: A review of the applicability of methods in practice. *Technological Forecasting and Social Change*, *171*, 120939. https://doi.org/10.1016/j.techfore.2021.120939

Wang, H., Liang, Q., Hancock, J. T., & Khoshgoftaar, T. M. (2024). Feature selection strategies: a comparative analysis of SHAP-value and importance-based methods. *Journal of Big Data*, *11*(1). https://doi.org/10.1186/s40537-024-00905-w

Wang, S., Yang, P., Tan, Q., Yao, L., & Li, C. (2025). Identifying spatially targeted organic fertilizer subsidies for water pollution control: An interactive hydro-economic model. *Water Research*, *281*, 123662. https://doi.org/10.1016/j.watres.2025.123662

Wu, P., Huang, J., Zheng, Y., Yang, Y., Zhang, Y., He, F., Chen, H., Quan, G., Yan, J., Li, T., & Gao, B. (2019). Environmental occurrences, fate, and impacts of microplastics. *Ecotoxicology and Environmental Safety*, *184*, 109612. https://doi.org/10.1016/j.ecoenv.2019.109612

Yin, H., Qi, Q., Wu, Y., Yan, J., & Shi, R. (2025). Assessment and early warning of the coupled carrying capacity of water resources, environment, and ecology: A case study of Ordos. *Ecological Indicators*, *180*, 114346. https://doi.org/10.1016/j.ecolind.2025.114346

Zhang, W., Wu, J., Yao, H., & Wang, J. (2025). Influence of drought identification methods on analyzing and assessing responses of water quality to droughts. *Ecological Indicators*, *181*, 114458. https://doi.org/10.1016/j.ecolind.2025.114458

Zhou, W., Yang, Y., Han, W., Huang, A., & Han, Q. (2025). Temperature-gradient drives shifts in macrobenthic communities near nuclear power plant thermal discharges: A case study from the Sanmen bay. *Marine Environmental Research*, *213*, 107670. https://doi.org/10.1016/j.marenvres.2025.107670

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- TABLE.docx